DOCUMENT RESUME

ED 232 569 HE 016 493

AUTHOR Clark, Sheldon B.; Nichols, James O.
TITLE Increasing the Precision of Estimates in Follow-Up Surveys: A Case Study. AIR 1983 Annual Forum Paper.
PUB DATE May 83
NOTE 17p.; Paper presented at the Annual Forum of the Association for Institutional Research (Toronto, Ontario, May 23-26, 1983).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; *Followup Studies; *Graduate Surveys; Higher Education; Institutional Research; *Predictive Measurement; *Research Problems; *Sampling; Statistical Analysis; *Surveys
IDENTIFIERS *AIR Forum

ABSTRACT

Survey data concerning teacher education program graduates were used to demonstrate the advantages of a stratified random sampling approach, with followup, relative to a one-shot mailing to an entire population. Sampling issues involved in such an approach are addressed, particularly with regard to quantifying the effects of nonresponse on the results. Comparisons were made between the two procedures, using the precision of estimates of population characteristics as the criterion. Much narrower confidence bounds were associated with the stratified random sampling approach, even though a smaller number of questionnaires had been returned. It is concluded that a redirection of energies from reaching the entire population to increasing the rate of response in a subset of that population has a significant payoff in terms of the precision of the estimates that can be made from the data. Keeping in mind that this advantage will be even greater in cases in which response variation differs across strata, the arguments are strong for adopting a stratified random sampling approach in situations such as the followup of teacher education program graduates. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED232569

#13

Increasing the Precision of Estimates in Follow-Up Surveys: A Case Study

Sheldon B. Clark, Assistant Director
James O. Nichols, Director

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

University Planning and Institutional Research
205 Lyceum
University of Mississippi
University, Mississippi 38677
(601) 232-7387

NE 016 493

Running head: Increasing the Precision of Estimates

THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Twenty-Third Annual Forum of the Association for Institutional Research held at the Sheraton Centre in Toronto, Ontario, May 23-26, 1983. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of Forum papers.

D. R. Coleman, Chairman
Forum Publication
Advisory Committee

Abstract

The authors used data gathered in a survey of graduates of a teacher education program to demonstrate the superiority of a stratified random sampling approach, with follow-up, relative to a one-shot mailing to an entire population. Sampling issues involved in such an approach have been considered, particularly with regard to quantifying the effects of nonresponse on the results. Comparisons were made between the two procedures, using the precision of estimates of population characteristics as the criterion. Much narrower confidence bounds were associated with the stratified random sampling approach, even though a smaller number of questionnaires had been returned.

Increasing the Precision of Estimates in Follow-Up Surveys: A Case Study

One of the tasks commonly charged to research wings of colleges of education is the assessment of the quality of teacher education programs. Among the several strategies employed to address this responsibility at a particular university is a follow-up survey of recent graduates. Historically, this has been accomplished by mailing questionnaires to all graduates of the program for a given year and making inferences using the questionnaires returned. Since the number of graduates of this program is substantial (about 600 in a recent year), the costs associated with the printing, mailing, handling, and postage for this single mailing have long been barriers to the inclusion of well-proved methods for increasing response rate (e.g., preletters, reminders). As a consequence of the single mailing, the return rate seldom exceeded 48%. Inferences based upon such a return are questionable, at best. During the past four years, however, a different approach has been taken which addresses the justifiable concerns about low response, while holding costs at roughly the same level. It is hoped that this case study will encourage others to consider adopting a similar approach.

Sampling Issues

One of the intuitively appealing aspects of sending questionnaires to an entire population is the knowledge that everyone

(excluding those who fail to receive the questionnaire through no fault of the evaluator) has the opportunity to voice his/her opinion. Also, the total number of responses generally exceeds that of a survey sample, even when several reminders are sent to nonrespondents in the sample. An overriding concern of experts in the field, however, is how those who do not respond differ from those who do respond (Brownlee, 1957; Bryson, 1976; Cochrane, 1977; Donald, 1960; Hartley & Rao, 1969; Kish, 1965; Lansing, Ginsburg, & Braaten, 1961; Moser, 1972). Much of the empirical research done in the area uses data collected from those who eventually respond, but adds time (elapsed time from mailing of questionnaire to receipt of completed questionnaire) as a variable. Other approaches have employed personal contact to interview those who do not return questionnaires by mail. The problem with these approaches to assessing the effect of nonresponse lies in the increased likelihood of what Sudman (1976) calls "response effects [,] which are the differences between reported and 'true' measures of behavior, characteristics, or attitudes" (p. 17). In addition to the problem that the characteristics of true nonrespondents are never measured, some attempts to quantify these characteristics are mathematically complex (Neyman, 1969; pp. 712-732); less complex methods are very conservative and limited in applicability (Cochrane, 1977, pp. 359-363). The concensus among experts, however, is that nonresponse is an important issue with far-reaching implications for the evaluator.

What, then, constitutes an acceptable return rate for surveys?

Although almost as many different recommendations exist as there are those who make them, most authors agree that the purpose of the survey should determine the acceptable rate. As Babbie (1973) points out, however, "there are some rules of thumb that might be followed. . . A response rate of at least 50 percent is adequate for analysis and reporting. A response rate of at least 60 percent is good, and a response rate of 70 percent is very good" (p. 165). Since the 48% response rate in the present context only approaches the lower limit of adequacy, the question of whether or not to change the method of surveying graduates becomes a matter of weighing the advantages to be realized by focusing on a sample of the population, rather than the entire population. A realistic benefit of the former approach is that several follow-ups (i.e., reminders) of nonrespondents can be included in the survey process for about the same expenditure (in terms of effort and cost) required to send and process a single mailing to the entire population.

Sampling Procedure

One of the strongest statements in favor of sampling comes from Williams (1978): "Samples almost always are more accurate" (p. 47). The basis for this conclusion is evidence that the percentage of nonsampling errors (i.e., nonrandom errors not asscciated with the sample itself, such as keypunching errors) increases as the size of the data set increases (p. 46).

Once the decision to draw a sample has been made, an appropriate sampling technique is the next consideration. A technique

applicable to the present situation is suggested by Cochrane (1977). He states that, "if intelligently used, stratification nearly always results in a smaller variance for the estimated mean or total than is given by a comparable simple random sample" (p. 99). Sudman (1976) makes the same recommendation from a different perspective:

> Given these constraints [(fixed amount of money to be spent, methods and costs of data collection known)], the efficiency of a sample is measured by the size of its sampling error relative to other samples that cost the same. Stratified sampling is intended to provide the smallest sampling error and hence the most information for the available resources. (p. 107)

Some authors claim that stratification also improves the randomness of the sample (Blalock, 1979, p. 561), while others disagree (Sudman, 1976, p. 108). This issue aside, the fact that there are significant gains to be made by using stratified random sampling procedures (in terms of the precision of the estimates), should the population strata indeed be heterogeneous with respect to response variance, together with the knowledge that no loss of precision will result, should the population strata be homogeneous, make convincing arguments for the use of stratified random sampling procedures. The likely characteristics of the population of teacher education graduates lend themselves well to stratification: the natural division of graduates by area of emphasis or department probably meet Cochrane's criterion of "intelligently used stratification." Let us now

examine the specific application and see how stratification with follow-up affects the precision of the estimates of population characteristics.

Comparison Between Stratified Random Sampling, With Follow-Up, and One-Shot Mailing to Entire Population

The breakdown--by department--of graduates of the teacher education program at this university in a recent year is listed in Table 1.

Insert Table 1 about here

Since the questionnaire consists of some categorical-response, demographic items (e.g., area of certification, whether or not employed as a teacher) and some Likert-response, evaluative items (e.g., degree of adequacy of training in classroom management), a desirable goal might be to estimate the percentages of the population that fall into each response category.

As in most other areas of statistics when estimates of multiple parameters are desired, a conservative estimate of the sample size needed to estimate a single parameter with specified precision is calculated by considering the worst possible case. By doing so, the reseacher is assured that all other parameters will be estimated with at least as much precision. In the case of estimating proportions, the worst possible case occurs when the likelihood of a person's being in one of two categories is equal (i.e., 50%).

Using the conventional 5% maximum tolerable error as a criterion,

the total sample size required in order to achieve this degree of precision can be calculated as follows (using the notation of Scheaffer, Mendenhall, & Ott, 1979):

$$n = \left[\sum_{i=1}^{L} (N_i^2 \, p_i q_i \, / \, W_i)\right] + \left[N^2 \, (B^2/4) + \sum_{i=1}^{L} N_i \, p_i q_i\right],$$

where n = total sample size required

i = index of stratum

L = number of strata

N_i = total number of subjects in stratum i

N = total number of subjects in all L strata

$p_i = q_i = .50$ = probability of an individual's responding or not responding in a particular manner (worst possible case)

$W_i = N_i/N$ = proportion of subjects in stratum i

$B = .05$ = maximum tolerable error.

Using this scheme, the number of subjects sampled from each stratum listed in Table 1 was 18, 55, 75, 22, 46, and 24, respectively, for a total sample size of 240. If all 240 individuals had returned completed questionnaires, the maximum error rate for any one item would have been 5%. In reality, however, only 80% (192) of the original sample returned the questionnaires after four reminders had been mailed. The proportion of nonresponses was approximately equal across strata. For the purpose of this discussion, let us assume that response variances were essentially homogeneous across strata (the case in which stratification would serve the

least purpose). Using the previous 48% response rate--obtained when the entire population was surveyed (which would have yielded 288 completed questionnaires)--let us compare the precision of the parameter estimates. An extension of a procedure outlined by Cochrane (1977, p. 361) allows us to do so.

Consider the 240 subjects identified in the stratified random sample to be a population--the target population, in effect, which would yield a maximum error rate of 5% in estimating the parameters associated with all 599 individuals. This population consists of two groups--those who responded (N_r) and those who did not respond (N_{nr}). If there are N subjects in this population ($N = 240$), the proportion of respondents is $N_r/N = W_r$, and the proportion of nonrespondents is $N_{nr}/N = W_{nr}$. Let P represent the actual proportion of all 240 subjects who fall into a particular response category, P_r represent the proportion of respondents who fall into that response category, and P_{nr} represent the proportion of nonrespondents who fall into that category. Since $P = W_r P_r + W_{nr} P_{nr}$, the bias in P_r caused by the nonrespondents, is $P_r - P$, or $P_r - (W_r P_r + W_{nr} P_{nr})$.

Algebraically, $P_r - (W_r P_r + W_{nr} P_{nr}) =$

$P_r = W_r P_r - W_{nr} P_{nr} =$

$P_r (1 - W_r) - W_{nr} P_{nr}.$

But, $1 - W_r = W_{nr}$,

so $P_r W_{nr} - W_{nr} P_{nr} =$

$W_{nr} (P_r - P_{nr}) =$ the bias in P_r.

Since the unknown proportion of P_{nr} must lie between 0% and 100%, and, since W_{nr} is known, confidence limits can be constructed around the population proportion P using the traditional concept of the standard error of measurement. Ninety-five percent confidence bounds for P_r can be expressed as follows:

$$\dot{p}_r \pm 2 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{\frac{1}{2}},$$

where $\dot{p}_r$ represents the sample proportion. In order to construct confidence bounds around P, it is ultimately conservative to assume that $P_{nr} = 0$ when finding the lower bound and $P_r = 1$ when finding the upper bound, since the two events are mutually exclusive:

$$\hat{P}_{\text{Lower}} = W_r \left\{ \dot{p}_r - 2 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{\frac{1}{2}} \right\} + W_{nr} (0) \text{ and}$$

$$\hat{P}_{\text{Upper}} = W_r \left\{ \dot{p}_r + 2 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{\frac{1}{2}} \right\} + W_{nr} (1).$$

Thus, $\text{Prob} (\hat{P}_{\text{Lower}} \leq P \leq \hat{P}_{\text{Upper}}) > .95$.

Using the rules governing joint probabilities, the initial 95% confidence (relative to the population parameter) associated with the calculated sample size of 240, together with the 95% confidence (relative to the sample statistic) associated with the 80% response rate, yield an overall confidence of 90.25% in the stratified random sampling procedure. In order to compare the relative precision of parameter estimates of the two procedures, 90.25% confidence limits were calculated for the hypothetical 48% population-based survey as well (see Figure 1 for calculations).

Insert Figure 1 about here

If a parameter estimate of .50 is calculated, for example, the confidence bounds are as follows:

CASE 1 (N = 599; 80% return rate of stratified random sample of 240 = 192)

$$\text{Prob } (.3742 \leq P \leq .6258) > .9025$$

CASE 2 (N = 599; 48% return rate = 288)

$$\text{Prob } (.2268 \leq P \leq .7732) > .9025 .$$

Discussion

Considering the relative widths of the confidence bounds for the population-based survey approach (Case 2) and the stratified random sampling approach (Case 1), .55 and .25, respectively, it is clear that the redirection of energies from reaching the entire population to increasing the rate of response in a subset of that population has a significant payoff in terms of the precision of the estimates which can be made from the data. Keeping in mind that this advantage will be even greater in cases in which response variation differs across strata (a likely situation), the arguments for adopting a stratified random sampling approach in situations such as the follow-up of teacher education graduates, are compelling.

References

Babbie, E. R. Survey research methods. Belmont, California: Wadsworth Publishing, 1973.

Blalock, H. M., Jr. Social statistics (Rev. ed.). New York: McGraw-Hill, 1979.

Brownlee, K. A. A note on the effects of nonresponse on surveys. Journal of the American Statistical Association, 1957, 52(277), 29-32.

Bryson, M. C. The Liberary Digest poll: making of a statistical myth. The American Statistician, 1976, 30(4), 184-185.

Cochrane, W. G. Sampling techniques (3rd ed.). New York: John Wiley & Sons, 1977.

Donald, M. N. Implications of nonresponse for the interpretation of mail questionnaire data. Public Opinion Quarterly, 1960, 1, 99-114.

Hartley, H. O. & Rao, J. N. K. A new estimation theory for sample surveys, II. In N. L. Johnson & H. Smith, Jr. (Eds.), New developments in survey sampling. New York: Wiley-Interscience, 1969.

Kish, L. Survey sampling. New York: John Wiley & Sons, 1965.

Lansing, J. B., Ginsburg, G. P., & Braaten, K. An investigation of response error. Urbana, Illinois: Bureau of Economic and Business Research, University of Illinois, 1961.

Moser, C. A. & Kalton, G. Survey methods in social investigation. New York: Basic Books, 1972.

Neyman, J. Bias in surveys due to nonresponse. In N. L. Johnson & H. Smith (Eds.), New developments in survey sampling. New York: Wiley-Interscience, 1969.

Scheaffer, R. L., Mendenhall, W., & Ott, L. Elementary survey sampling (2nd ed.). North Scituate, Massachusetts: Duxbury Press, 1979.

Sudman, S. Applied sampling. New York: Academic Press, 1976.

Sudman, S., & Bradburn, N. M. Asking Questions. San Francisco: Jossey-Bass, 1982.

Williams, B. A sampler on sampling. New York: John Wiley & Sons, 1978.

Table 1

Number of Teacher Education Graduates, by Department

Department	Number of Graduates
Art and Music Education	46
Curriculum and Instruction	136
Elementary Education	187
Health and Physical Education	55
Special Education	115
Vocational-Technical Education	60
TOTAL	599

Figure 1

Computation of Confidence Bounds

Stratified Random Sample

$$W_r = N_r/N = 192/240 = .80$$

$$W_{nr} = N_{nr}/N = 48/240 = .20$$

$$\hat{P}_L = W_r \left\{ \dot{p}_r - 2 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{1/2} \right\} + W_{nr}(0)$$

$$= .80 \left\{ .50 - 2 \left[(.25/192)(1 - 192/240) \right]^{1/2} \right\} + 0$$

$$= .80 \left[.50 - 2(.0161) \right]$$

$$= .3742$$

$$\hat{P}_U = W_r \left\{ \dot{p}_r + 2 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{1/2} \right\} + W_{nr}(1)$$

$$= .80 \left[.50 + 2(.0161) \right] + .20(1)$$

$$= .6258$$

Population-Based Survey $[z(.9025) = 1.30]$

$$W_r = N_r/N = 288/599 = .48$$

$$W_{nr} = N_{nr}/N = 311/599 = .52$$

$$\hat{P}_L = W_r \left\{ \dot{p}_r - 1.3 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{1/2} \right\} + W_{nr}(0)$$

$$= .48 \left\{ .50 - 1.3 \left[(.25/288)(1 - 288/599) \right]^{1/2} \right\} + 0$$

$$= .48 \left[.50 - 1.3(.0212) \right]$$

$$= .2268$$

$$\hat{P}_U = W_r \left\{ \dot{p}_r + 1.3 \left[(p_r q_r / N_r)(1 - N_r/N) \right]^{1/2} \right\} + W_{nr}(1)$$

$$= .48 \left[.50 + 1.3(.0212) \right] + .52(1)$$

$$= .7732$$